Interactive Machine Learning
MAS.S62

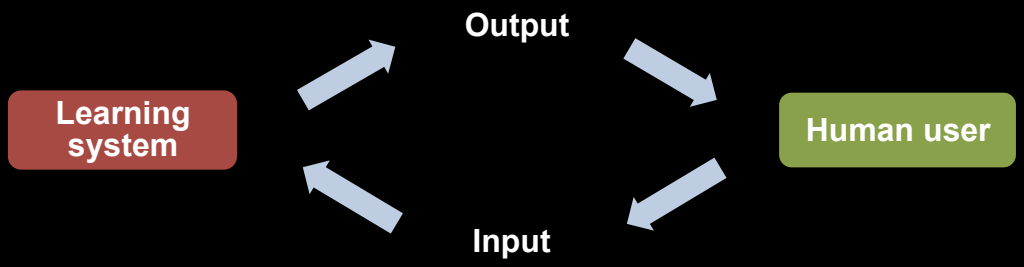
Brad Knox: the brief story

- Undergrad: psychology, pre-med, and a little philosophy at Texas A&M
- The year after: quarter-life crisis
- Post-bacc and Robocup at UT Austin
- PhD with Peter Stone at UT Austin
  - Dissertation: *Learning from human-generated reward*
- Postdoc in MIT Media Lab with Cynthia Breazeal

Interactive Machine Learning (IML)

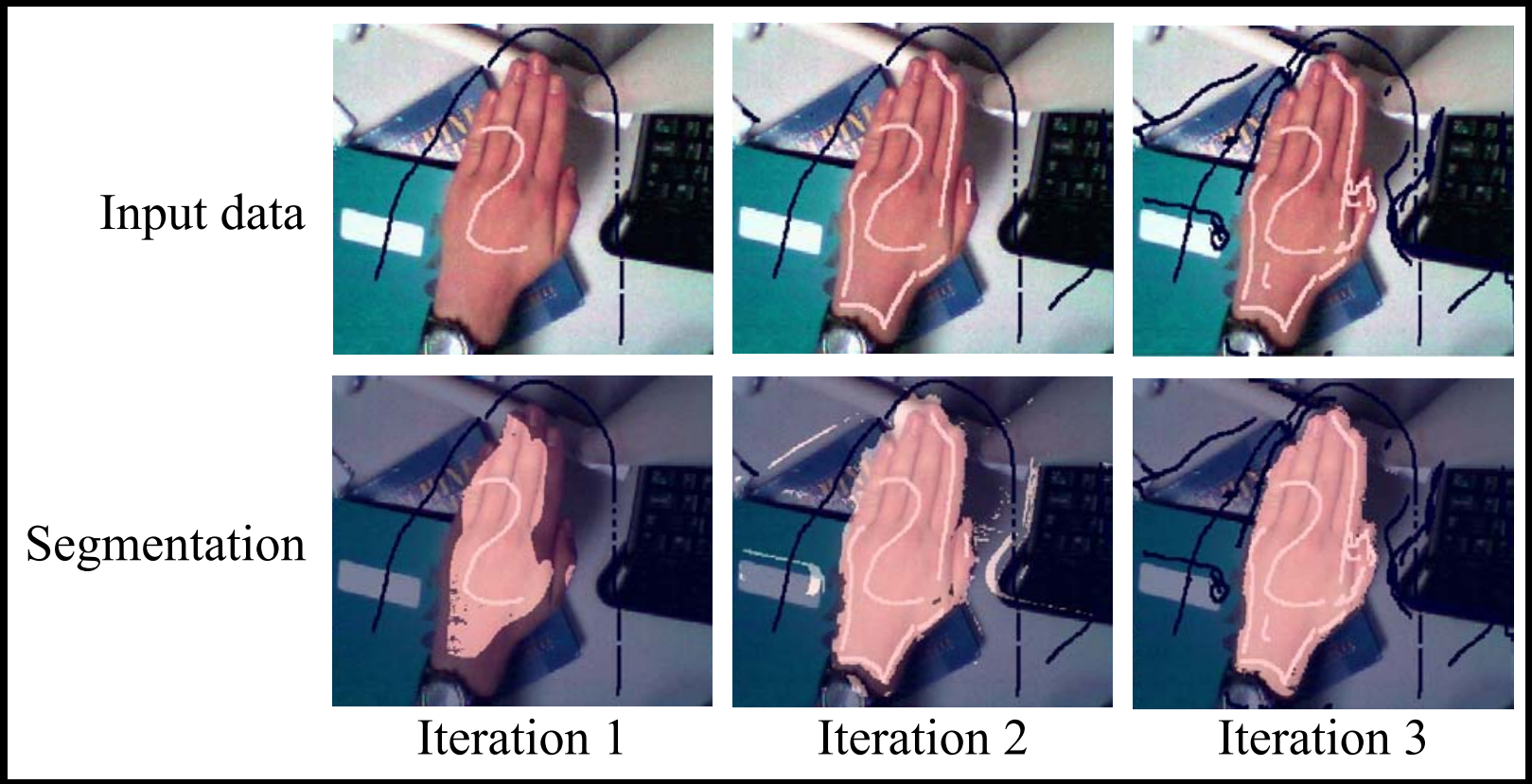
Course definition - machine learning such that

1. a human is in a tight machine learning loop, observing the result of learning and providing input that affects further learning, and
2. the human intentionally provides input, fulfilling a teaching role.



IML examples

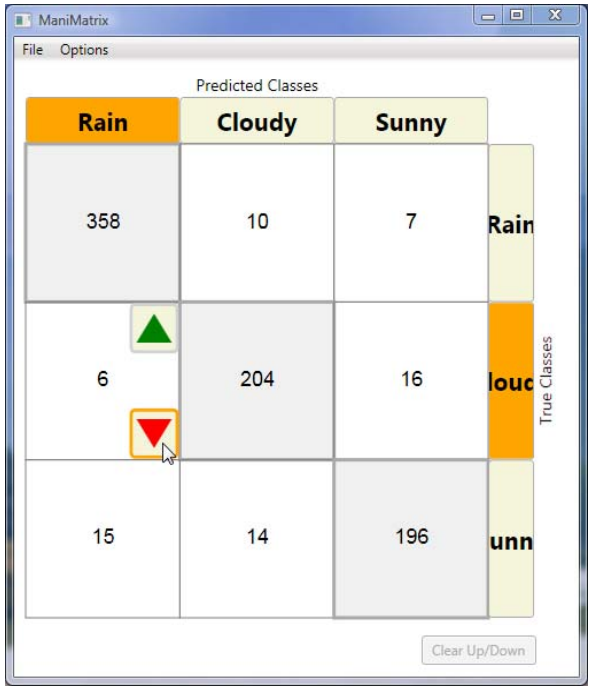
Interactive image segmentation

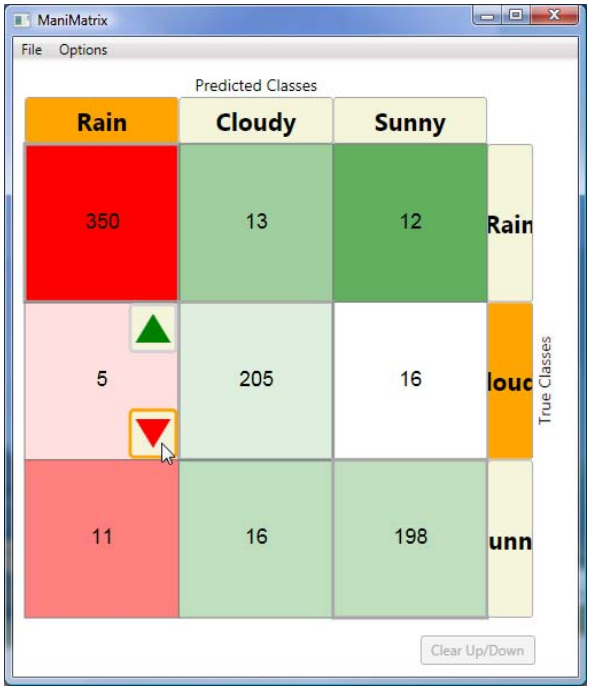


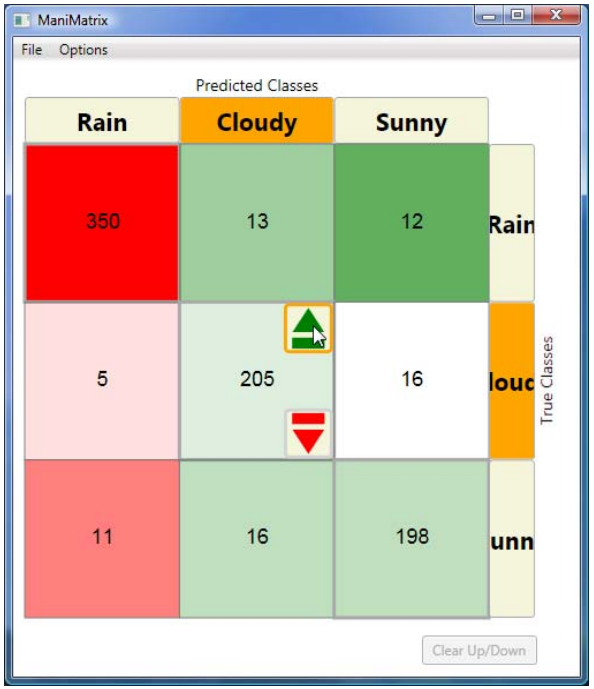
Fails and Olsen (2003)

IML examples

ManiMatrix – user can manipulate confusion matrix, affecting the cost of different errors



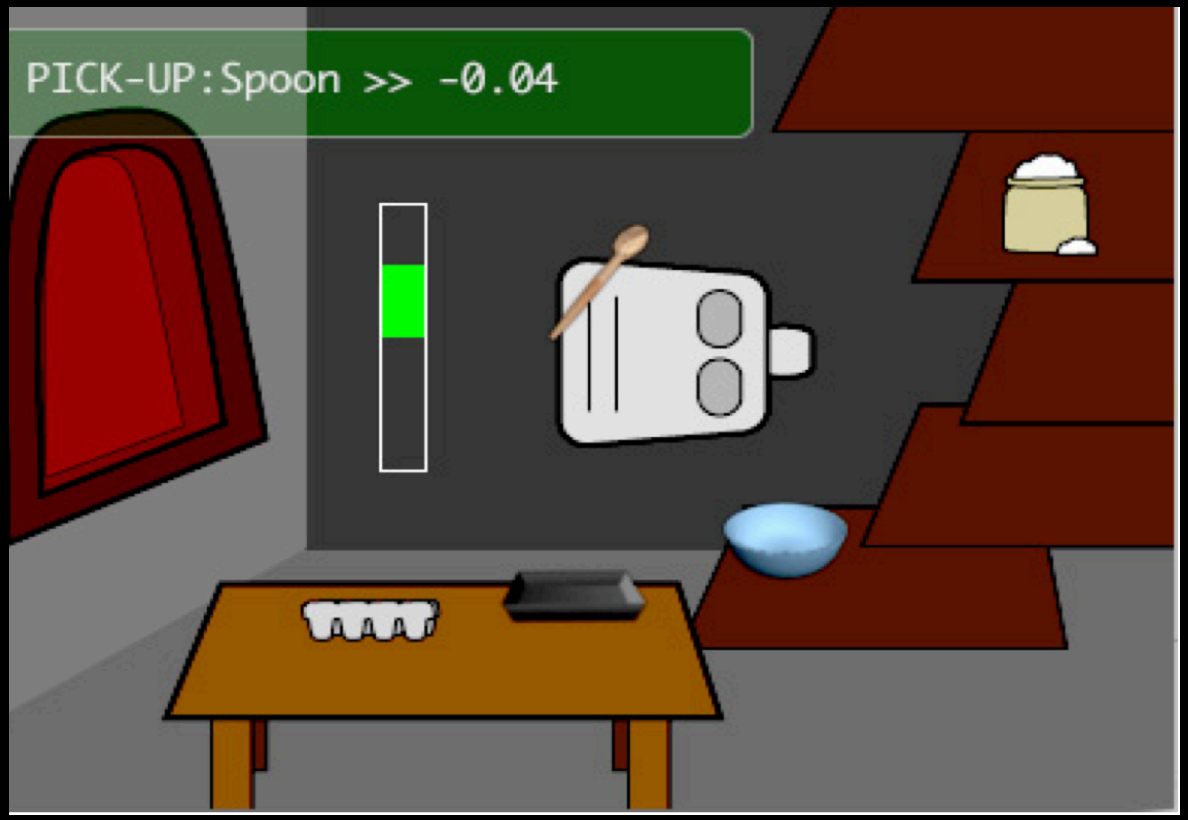




Kapoor et al. (2010)

IML examples

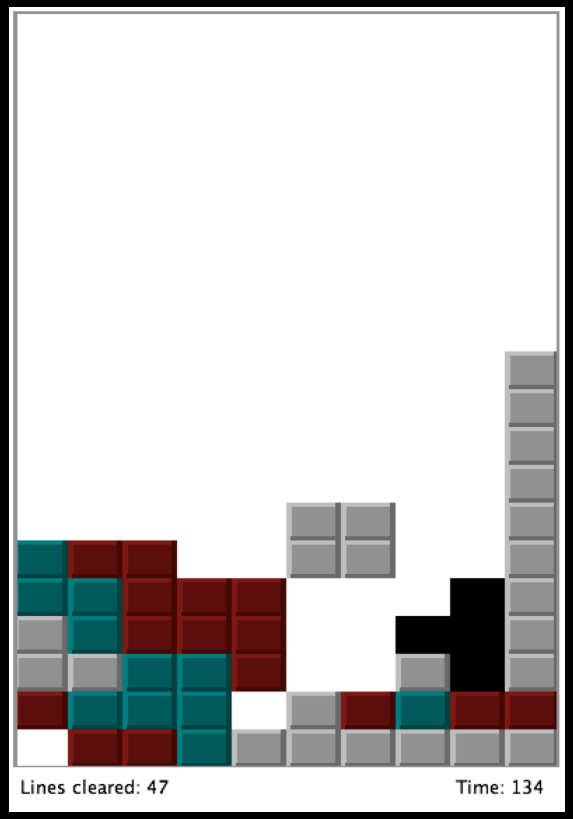
Interactive reinforcement learning

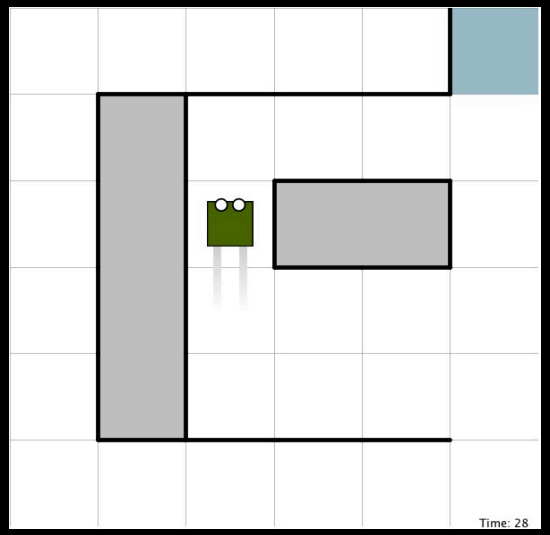


Thomaz and Breazeal (2008)

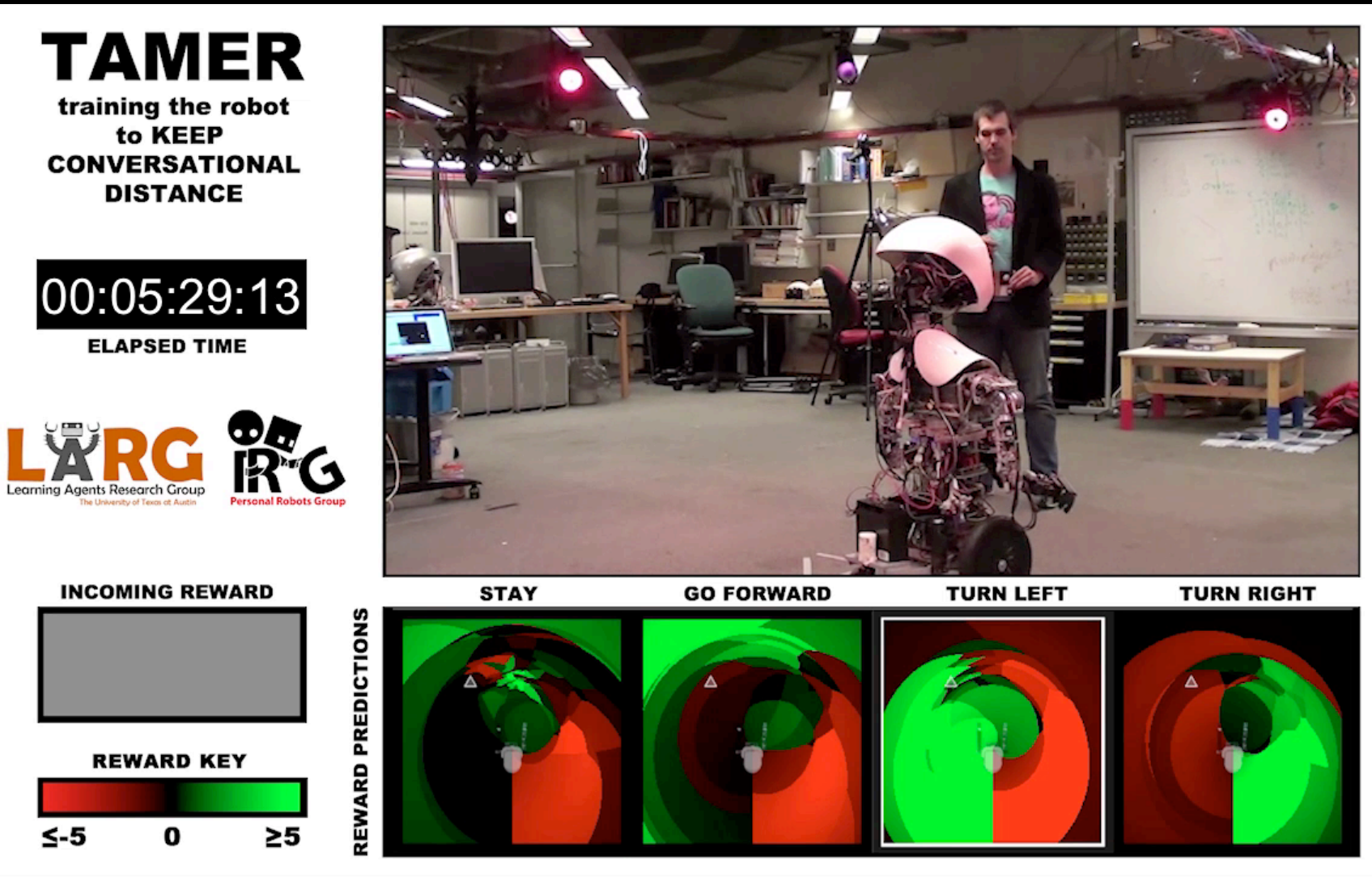
IML examples

Interactive reinforcement learning (demo)





Knox and Stone (2009, 2012)



Knox, Breazeal, and Stone (2013)

IML examples

Interactive training of musical instruments (demo)

Fiebrink et al. (2009)

IML in the wild

- Netflix's recommender system
- Pandora's recommender system
- [Borderline] Spam filters that allow you to label and unlabel spam

What's not IML?

Human detection for surveillance

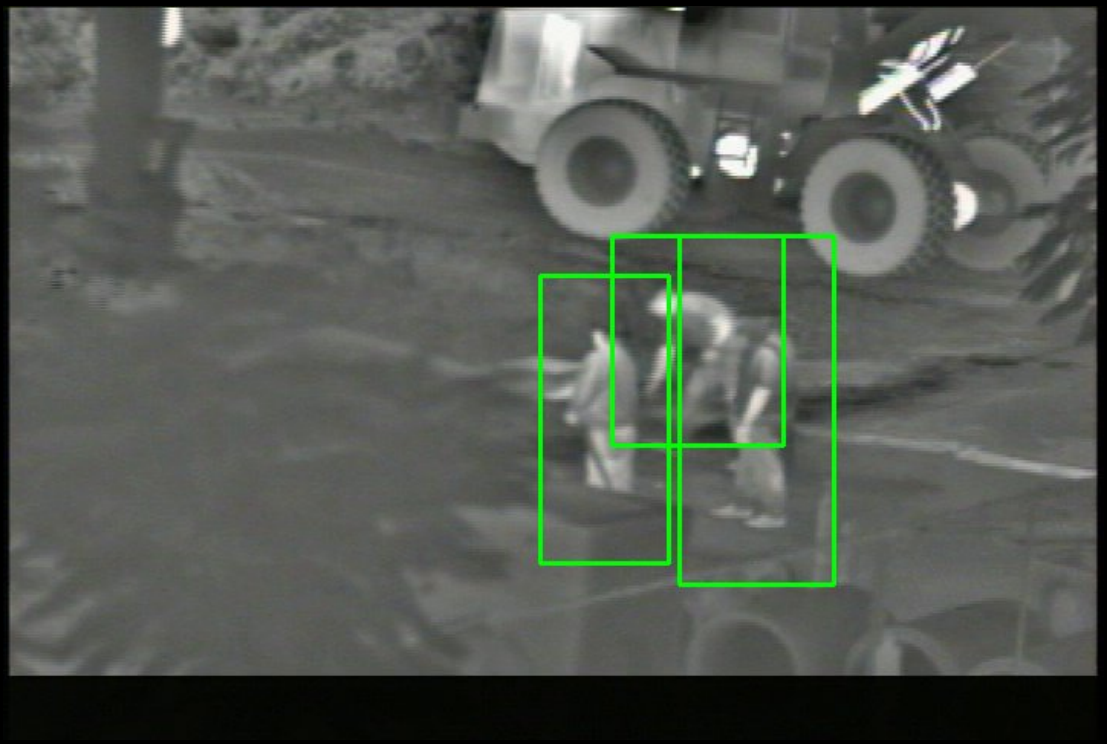


Photo from Bo Wang's (USC) website

Possibly interactive, but no teaching role

What's not IML?

Credit card fraud detection

Possibly interactive and may have a teaching role, but no tight loop

What's not IML?

A reinforcement learning agent interacting with its environment under typical circumstances.

If there's a human, he or she is generally not in a teaching role.

What's not IML?

Clever and interactive labeling systems

Von Ahn (2006)

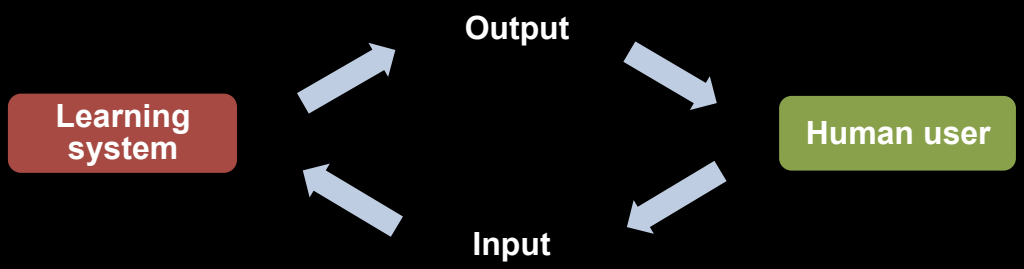
Interaction is meant to teach a learner, but interaction occurs *before* learning.

Interactive Machine Learning (IML)

Course definition *revisited and simplified*

ML such that

1. a human is in a tight machine learning loop
2. the human's input is meant to teach



Why does IML deserve its own focus?

With only ML expertise, the human becomes a black box.

- Unrealistic assumptions about the human are frequently made

With only HCI expertise, the machine learning algorithm becomes a black box.

- Opportunities for interaction channels are missed or simply can't be implemented.
- "I tried it and it doesn't work."

Why does IML deserve its own focus?

Much of applied machine learning is interactive anyway.

We should understand and design for this interaction.

Who is here

This course: *resources*

website with syllabus and schedule

https://stellar.mit.edu/S/course/MAS/fa13/MAS.S62/index.html

This course: *progression of topics*

1. Supervised learning
2. Evaluation by techniques/standards of
   - Machine learning
   - Human-computer interaction
3. (Maybe) unsupervised learning (e.g. clustering)
4. Sequential decision making
   - Learning from demonstration
   - Reinforcement learning

(We will be looking at the interactive versions of each of these ML methods.)

This course: *goals*

- Understand the current literature on IML
- Identify unexplored research topics
- Practice finding and choosing research problems
- Become acquainted with sequential decision-making and research methods in ML and HCI

This course: *goals*

- Identify common themes and goals within IML research
  - especially those unique to the intersection of HCI and ML
- Hands-on experience creating interactive machine learning systems
- Conduct novel research on IML!

This course: *course components*

- Limited pedagogical lectures
- Guest lectures
- Readings with responses posted and class discussion
- Preliminary research assignments
  - Toolkit presentations
  - Small hands-on projects
- Final research project

This course: *confirmed guest lecturers*

Nick Gillian (MIT)
Rebecca Fiebrink (Princeton)
Simone Stumpf (City University London)
Kayur Patel (Google)
Krzysztof Gajos (Harvard)
Ashish Kapoor (Microsoft Research)
Joe Konstan (Univ. of Minnesota)
Henry Lieberman (MIT)

Sitting in

People are welcome to sit in, but they must *keep up with the reading and submit reading responses* to join on reading discussion days.

To join for lectures, student presentations, and guest lectures, no preparation is required.

First reading assignment

Two readings and a response:

A Few Useful Things to Know about Machine Learning by Domingos

and

Machine Learning that Matters by Wagstaff

Reading responses: What to write

Full instructions will be posted by tomorrow night on course website.

Mostly free-form, with credit based on evidence that you have done the readings carefully.

~1 page, single-spaced, 12pt Times font

Reading responses: What to write

Acceptable responses include:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the reading to previous readings;
- Solutions to problems or exercises posed in the readings;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the paper's importance; and
- Summaries of the most important things you learned.

Reading responses: What to write

There is one specific requirement though, creating *a discussion point* that you may be asked to bring up in class.

Possible discussion topics (note overlap with previous list):

- Controversial questions/answers
- Points of confusion
- Ideas about extensions to the work
- Insights on broader topics that are relevant to this work